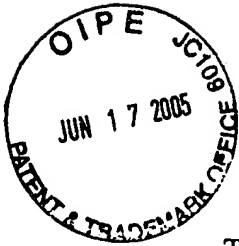


# METHOD AND APPARATUS FOR CONTENT MANIPULATION



## Statement of the Invention

This invention relates to method and apparatus for manipulating data content from one state to another. For example, one embodiment of the present invention is an apparatus for translating words and phrases from one language to another language. In this example, the data content (words and phrases) are manipulated (translated) from one state (a first language) to a second state (a second language). More specifically, the invention relates to a technique whereby data content is parsed and manipulated so as to provide an accurate conversion into another state.

## BACKGROUND OF THE INVENTION AND PRIOR ART

This invention relates to the field of data manipulation. The manipulation process is dynamic – the manipulation becomes faster and more accurate the more data is input and converted into a second state. Moreover, the input of data that is manipulated can be automatic – upon the creation of initial parameters, the method allows the ready manipulation from one state to another. For example, in the embodiment of the present invention that focuses on language translation, the initial parameters may be established by having a computer receive the same data in three different languages (i.e., the same book translated into three different languages). Once these initial parameters are established, the method will utilize a database that relates words and phrases in any one language to identical words and phrases in the other two languages. In addition, the

method utilizes the placement and context of words within sentences and phrases (i.e., component parts of a structure representing a specific concept) to extrapolate the inventive manipulation method.

In particular, a preferred embodiment of the present invention makes it possible to allow a computer to receive data input in one language, translate that data, and provide a data output in a separate and distinct second language. The embodiment of a translation method utilizes an overlapping translation technique to obtain an accurate translation of words, phrases, and sentences. Prior art translation devices, particularly those involved utilizing a computer, focus on individual word translation, or utilize special codes to facilitate the ready translation from a first language into a second language. The present invention enables words and phrases in a first language to be accurately translated in their correct context in the exact manner those words and phrases would have been written in a second language.

## **BRIEF DESCRIPTION OF THE INVENTION**

According to one embodiment of the present invention, there is provided a language translation system that utilizes a database to accurately and automatically translate data from a first language to a second language. The languages can be any type of conversion and are not necessarily limited to spoken/written languages; for example, the conversion can encompass computer languages, specific data codes such as ASCII, and the like. The database is dynamic; i.e., the database grows as content is input into the translation system, with successive iterations of the translation system using content

entered at a previous time. The preferred embodiment of the invention utilizes a computing device such as a personal computer system of the type readily available in the prior art. However, the system does not need to use such a computing device and can readily be accomplished by other means, including manual creation of the database and translation methods. Moreover, the method of data input can be automatic – in the translation example, the data can be input with certain defined parameters that allows the computer to automatically recognize the translation from one language to another.

#### **DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT OF THE INVENTION**

A preferred embodiment of the present invention will now be described.

The present invention may be utilized on a common computer system having at least a display means, an input method, and output method, and a processor. The display means can be any of those readily available in the prior art, such as cathode ray terminals, liquid crystal displays, flat panel displays, and the like. The processor means also can be any of those readily available and used in a computing environment such that the means is supplied to allow the computer to operate to perform the present invention. Finally, an input method is utilized to allow the input of data in a first language, and an output method is utilized to allow the output of translated data in a second language.

For the purposes of describing the preferred embodiment, a system wherein data in the English language is translated to data in the Hebrew language; these selections are

for descriptive purposes only and are not meant to limit the selection of a first and second language.

According to a preferred embodiment of the present invention, the computer system operates to create a database of translations from English to Hebrew. The translation method encompasses at least the following steps:

First, data in the English language is input into the computer system. The computer system then operates a database to parse the input data into natural delimited segments, such as words.

Second, the parsed data is examined, by operation of a database, to determine if the parsed data is recognized as equivalent to data in the second language. That is, the database will include known word translations – if such translations are known, then the computer system, through operation of the database, will convert an English word into its Hebrew equivalent. If the translation is not included in the database, then the computer system will operate in a manner to query the user to input the appropriate translation. Thus, if the database does not know the Hebrew equivalent to an input English word, the computer will ask the user to provide the appropriate Hebrew equivalent. The user will then return the translation and input said translation into the database. Upon subsequent use, the computer system will operate the database in a manner such that the translation is known by virtue of its input by the user at an earlier point in time.

(Note that the database may be initially created by inputting known translations.

Thus, for example, the database may be initially created by inputting a list of English words and their Hebrew equivalents into the database prior to a first translation operation).

Thus, in a second step the input data is examined in its parsed state – e.g., word for word – and the appropriate translations are either return (by virtue of the operation of the database) or entered into the database.

Third, the input data is examined in a manner so as to increment the parsed segments. For example, if the data was first input on a word-by-word basis, the translation method of the present invention next examines the input data by evaluating phrases comprising two words. Thus, the entire input data is now translated, using the database, on a two-word phrase by phrase basis. Again, in a manner similar to that described above, the database returns translations for the two-word phrases if known; if unknown the translation system operates to query the user to input the appropriate translation for the two word phrase.

The incremented parsed segments are examined in an overlapping manner. For example, if four parsed segments exist (denominated 1, 2, 3, and 4), then the incremented parsed segment examination will first determine segments 1 and 2 combined, then determine segments 2 and 3 combined, then determine segments 3 and 4 combined.

Fourth, the system operates in a manner to combine the overlapped segments so as to eliminate redundant translated segments and provide a coherent result.

The above steps are reiterated out from 1 to an infinite number of steps (n) so as to provide the appropriate translation. The translation system works automatically by verifying consistent strings that bridge encoded word-blocks in both Languages. These automatic approvals for overlap-bridges that are consistent across both languages provide a language network that translates between two languages with perfect accuracy once the database reaches critical mass.

EXAMPLE: As an example, consider the English language phrase "I want to buy a car." Upon operation of a method of the present invention, this phrase will be input into a computer operating a database. The computer will operate to determine if the database includes Hebrew equivalents to the following words: "I", "want", "to", "buy", "a", and "car". If such equivalents are known, the computer will return the Hebrew equivalents. If such equivalents are not known, the computer will query the user to provide the appropriate Hebrew translations, and store such translations for future use.

Next, the computer will parse the sentence into two word segments in an overlapping manner: "I want", "want to", "to buy", "buy a" and "a car". The computer will operate to return the Hebrew equivalents of these phrase segments (i.e., the Hebrew equivalent if "I want" etc.); if such Hebrew equivalents are not known then the computer

will query the user to provide the appropriate Hebrew translations, and store such translations for future use.

The system continues in an incremented fashion until the entire phrase is input. That is, the system will next examine three-word segments "I want to", "want to buy", "to buy a", and "buy a car"; then, after the appropriate translation attempts, will proceed until the segment being examined is one phrase (in this case, the entire phrase "I want to buy a car").

The system, after going through this parsing, then compares the returned translation equivalents, eliminates redundancies, and outputs the translated phrase to the user.

As another example, consider the following sentence entered in English and intended to be translated into Hebrew: "In addition to my need to be loved by all the girls in town, I always wanted to be known as the best player to ever play on the New York state basketball team."

Through the process described above, the system will determine that the phrase "In addition to my need to be loved by all the girls" translates in Hebrew to "benosaf ltzorech sheli lihiot ahuv al yeday kol habahurot".

The system will also determine that the following phrase translations using the method described above:

“loved by all the girls in town” translates to “ahuv al yeday kol habahurot buir”;

“the girls in town, I always wanted to be known” translates to

“Habahurot buir, tamid ratzity lihiot yahua”;

“I always wanted to be known as the best player” translates to

“tamid ratzity lihiot yahua bettor hasahkan hachi tov”;

“the best player to ever play on the New York state basketball team” translates to

“hasahkan hachi tov sh hay paam sihek bekvutzat hakadursal shel medinat new york”.

With these returns by the database, the system will operate in a manner to compare overlapping phrases and eliminate redundancies.

“In addition to my need to be loved by all the girls” translates to

“benosaf Itzorech sheli lihiot ahuv al yeday kol habahurot”; and “loved by all the girls in town” translates to “ahuv al yeday kol habahurot buir”.

Now, utilizing an overlap technique, the system will take the English segments “In addition to my need to be loved by all the girls” and “loved by all the girls in town” and will examine the returned Hebrew segments “benosaf Itzorech sheli lihiot ahuv al yeday kol habahurot” and “ahuv al yeday kol habahurot buir” and determine the overlap. In English, the phrases are:



[illegible]

**THE** **NEW** **YORK** **PUBLIC** **LIBRARY**

**THE** **NEW** **YORK** **PUBLIC** **LIBRARY**

[illegible][illegible][illegible][illegible]

**THE** **NEW** **YORK** **PUBLIC** **LIBRARY**

**THE** **NEW** **YORK** **PUBLIC** **LIBRARY**

[illegible][illegible][illegible][illegible]

Remove the overlap:

“In addition to my need to be loved by all the girls in town”

“the girls in town, I always wanted to be known”

To yield:

“In addition to my need to be loved by all the girls in town, I always wanted to be known”.

In Hebrew:

“benosaf ltzorech sheli lihiot ahuv al yeday kol habahurot buir”

“Habahurot buir, tamid ratzity lihiot yahua”

Yields:

“benosaf ltzorech sheli lihiot ahuv al yeday kol habahurot buir, tamid ratzity lihiot yahua”

Continuing the process:

“In addition to my need to be loved by all the girls in town, I always wanted to be known” and “I always wanted to be known as the best player”.

Hebrew translations returned by the database are:

“benosaf ltzorech sheli lihiot ahuv al yeday kol habahurot buir, tamid ratzity lihiot yahua”  
and “tamid ratzity lihiot yahua bettor hasahkan hachi tov”.

Removing the English overlap yields:

“In addition to my need to be loved by all the girls in town, I always wanted to be known as the best player”.

Removing the Hebrew overlap yields:

“benosaf ltzorech sheli lihiot ahuv al yeday kol habahurot buir, tamid ratzity lihiot yahua bettor hasahkan hachi tov”

Continuing the process:

Operating on the two English phrases “In addition to my need to be loved by all the girls in town, I always wanted to be known as the best player” and “the best player to ever play on the New York state basketball team”.

Operating on the corresponding Hebrew phrases “benosaf ltzorech sheli lihiot ahuv al yeday kol habahurot buir, tamid ratzity lihiot yahua bettor hasahkan hachi tov” and “hasahkan hachi tov sh hay paam sihek bekvutzat hakadursal shel medinat new york”.

Two English sets before overlap removal

In addition to my need to be loved by all the girls in town, I always wanted to be known as the best player

the best player to ever play on the new york state basketball team.

Removing the English overlap yields:

“In addition to my need to be loved by all the girls in town, I always wanted to be known as the best player to ever play on the New York state basketball team”.

Removing the Hebrew overlap yields:

“benosaf ltzorech sheli lihiot ahuv al yeday kol habahurot buir, tamid ratzity lihiot yahua bettor hasahkan hachi tov sh hay paam sihek bekvutzat hakadursal shel medinat new york”, which is the translation of the phrase desired to be translated.

Upon the completion of this process, the system operates to return the translated final phrase and output said phrase.

As will be understood by those skilled in the art, many changes in the apparatus and methods described above may be made by the skilled practitioner without departing from the spirit and scope of the invention.